PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 :
C12N 15/85, 15/12, 5/10, 9/10, C12Q 1/68, A61K 48/00

A1

(11) International Publication Number: WO 94/29468

(43) International Publication Date: 22 December 1994 (22.12.94)

(21) International Application Number: PCT/GB94/01251

(22) International Filing Date: 10 June 1994 (10.06.94)

(30) Priority Data:
9312070.7 11 June 1993 (11.06.93) US

(71) Applicant *(for all designated States except US)*: THROMBOSIS RESEARCH INSTITUTE [GB/GB]; 90 Fetter Lane, London EC4A 1EQ (GB).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: BLEASDALE, Ciao, Chang [CN/GB]; 45 Oaklands, Westham, Pevensey, Eastbourne BN24 5AW (GB). DEMOLIOU-MASON, Catherine [GB/GB]; 5 Wellesley Mansion, Edith Villas, London W14 9AH (GB). KAKKAR, Vijay, Vir [IN/GB]; 6 Aspen Copse, Bickley, Kent BR1 2NZ (GB).

(74) Agents: LINN, Samuel, Jonathan et al.; Mewburn Ellis, York House, 23 Kingsway, London WC2B 6HP (GB).

(81) Designated States: AM, AT, AU, BB, BG, BR, BY, CA, CH, CN, CZ, DE, DK, ES, FI, GB, GE, HU, JP, KE, KG, KP, KR, KZ, LK, LU, LV, MD, MG, MN, MW, NL, NO, NZ, PL, PT, RO, RU, SD, SE, SI, SK, TJ, TT, UA, US, UZ, VN, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published

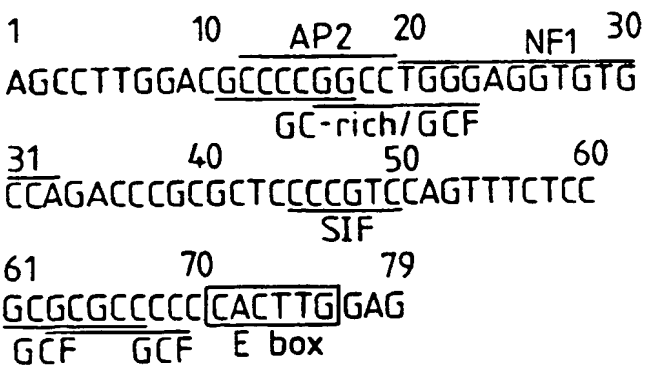
With international search report.

Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: UNTRANSLATED EXON-1 SEQUENCES OF EUKARYOTIC GENES AS PROMOTERS

(57) Abstract

Untranslated exon 1 sequences of certain eukaryotic genes, particularly SMHC and sTnI genes, without any 5' or 3' flanking sequences, have strong promoter activity in the expression of reporter genes such as CAT and β-gal in a variety of mammalian host cells, e.g. vascular smooth muscle cells. The invention provides methods of such enhanced gene expression, eukaryotic expression vectors therefor, and use of such promoters in the expression of such genes. The invention has particular applicability to the expression of genes for use in human gene therapy.



FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	**GB**	United Kingdom	**MR**	Mauritania
AU	Australia	**GE**	Georgia	**MW**	Malawi
BB	Barbados	**GN**	Guinea	**NE**	Niger
BE	Belgium	**GR**	Greece	**NL**	Netherlands
BF	Burkina Faso	**HU**	Hungary	**NO**	Norway
BG	Bulgaria	**IE**	Ireland	**NZ**	New Zealand
BJ	Benin	**IT**	Italy	**PL**	Poland
BR	Brazil	**JP**	Japan	**PT**	Portugal
BY	Belarus	**KE**	Kenya	**RO**	Romania
CA	Canada	**KG**	Kyrgystan	**RU**	Russian Federation
CF	Central African Republic	**KP**	Democratic People's Republic of Korea	**SD**	Sudan
CG	Congo			**SE**	Sweden
CH	Switzerland	**KR**	Republic of Korea	**SI**	Slovenia
CI	Côte d'Ivoire	**KZ**	Kazakhstan	**SK**	Slovakia
CM	Cameroon	**LI**	Liechtenstein	**SN**	Senegal
CN	China	**LK**	Sri Lanka	**TD**	Chad
CS	Czechoslovakia	**LU**	Luxembourg	**TG**	Togo
CZ	Czech Republic	**LV**	Latvia	**TJ**	Tajikistan
DE	Germany	**MC**	Monaco	**TT**	Trinidad and Tobago
DK	Denmark	**MD**	Republic of Moldova	**UA**	Ukraine
ES	Spain	**MG**	Madagascar	**US**	United States of America
FI	Finland	**ML**	Mali	**UZ**	Uzbekistan
FR	France	**MN**	Mongolia	**VN**	Viet Nam
GA	Gabon				

UNTRANSLATED EXON-1 SEQUENCES OF EUKARYOTIC GENES AS PROMOTERS

FIELD OF THE INVENTION

This invention relates to a novel function of untranslated TATA-less first exon sequences in eukaryotic genes, the use of these sequences in the expression regulation of eukaryotic genes, and for the construction of novel expression vectors which provide fine controlling of expression levels of specific genes *in vitro* and *in vivo*, for use for example in human gene therapy.

BACKGROUND OF THE INVENTION AND PRIOR ART

Background literature references are referred to herein by way of parenthetical numerical citation in the text to the appended bibliography. The disclosures of these references are incorporated herein by reference.

As is well known, all eukaryotic genes contain exons and introns, the multifunctions of which in gene regulation processes are still the subject of much investigation. A striking finding in recent progress of human genome mapping is the high degree of conservation of introns sequences, which predict the importance of these large regions whose functions are still largely unknown. Another interesting phenomenon is that in the case of many eukaryotic genes, their first exons are highly conserved, actively transcribed, but untranslated. Remarkable progress in the study of untranslated first exons has been made during recent years. For example, the first exons of a number of genes have been shown to play roles in transcription regulation; they are required either

for the full activity of their promoters, or for interaction with upstream cis-elements to regulate the expression level of genes. Those exonic control regions have been described for example in the case of the HSV *tk* gene where a portion of its first exon is required for full activity from the *tk* promoter (1), and in the case of skeletal troponin I (sTnI) gene, where the first exon is required for full muscle-specific activity of the sTnI promoter. The conclusion here is that the sTnI exon 1 does not contain a transcriptional enhancer, but the interaction between sTnI exon 1 and the distal upstream region is necessary for the expression regulatory mechanism (2). Other exonic control regions previously described include: RNA polymerase III transcribed rRNAs (3,4) and tRNAs (5), heat shock genes (6 – 9), the gastrin gene (10); c-myc is an interesting model, and shows it has at least four different promoters. The majority of transcripts start at the P_1 (-161) and P_2 promoter (+1), in normal Burkitt's lymphoma cells (BL), which account for about 10-25% and 75-90% of c-myc RNA. In BL cells with the chromosomal breakpoint upstream of the gene, c-myc is preferentially transcribed from the P_1 promoter, shifting the ratio of P_1/P_2 usage to >1. Notably the P2 promoter is located in the c-myc exon 1 region and contains a potential TATA box (29 – 23) which can be recognized by both polymerase II & III. The sequences within the P2 promoter affect the elongation and premature termination of transcripts initiated from the 5' upstream p1 promoter.

Two further c-myc promoters[3] are the P_0 promoter which is located 55-650 bp upstream of P_1 and the P_3 promoter which is located in the first intron of the gene. The potential coding capacity of these P_1/P_2, P_0 and P_3 RNAs is different. P_1/P_2 RNAs contain a single large open reading frame and encode two c-myc proteins of 64 and 67 kd translation of the 67 kd protein is initiated at a CTG codon at the end of exon 1, whereas

the 64 kd protein is initiated in exon 2, using an AUG codon. However, P_0 RNA contains three open reading frames, the 5' and middle open reading frames being initiated upstream of the P_1 promoter. The 3' open reading frame is identical to the P_1/P_2 – 64/67 kd proteins. P_3 RNA is initiated at multiple start sites within the first intron and lacks the exon 1 sequences, and is only able to code the 64 kd c–myc protein (11 – 14).

In eukaryotes, transcription is carried out by three different RNA polymerases: RNA polymerase I, II and III, each of which is dedicated to the transcription of different sets of genes. The genes in each class contain characteristic promoters, which usually consist of two types of functional elements: core (basal) promoter elements and modulator (upstream) promoter elements. The core promoter elements are sufficient to determine RNA polymerase specificity and direct low levels of transcription, whereas the modulator elements enhance or reduce the basal levels of transcription. The core promoter elements are first recognized by specific transcription factors, which then recruit the specific RNA polymerase. However recent studies have discovered that one general factor, the TATA box–binding protein (TBP), plays a role in all three polymerase systems, and the mechanisms of initiation complex assembly are strongly conserved. There are thus two fundamental mechanisms of initiation complex assembly: for TATA – containing promoters (polymerase II and III), TBP binding to TATA box directly
with its concave DNA–building surface, and its large convex surface to which other proteins i.e. activators, general factors, polymerase can bind: for non – TATA promoters (all three polymerases), another protein binds to the DNA, i.e. upstream binding factor for polymerase I family, a transcription factor like SP1 for polymerase II family, and

TFIIIC for classical polymerase III promoters. These proteins then recruit TBP directly or indirectly via a TBP–associated factor (TAF), once the TBP is in the initiation complex. It can then facilitate the binding of other proteins (15 – 17).

Transcription initiation is one of the most important ways to regulate gene expression. Although most protein–encoding genes contain a TATA motif, many promoters transcribed by polymerase II, especially those of housekeeping genes, lack a TATA element. These are referred to as TATA–less promoters. In some promoters, a second type of core promoter element called a transcriptional initiator (Inr) has been found to mediate the same functions as the TATA motif. An Inr can be defined as a DNA sequence element that overlaps a transcription start site and is sufficient for determining the start site location in a promoter that lacks a TATA box and also can cooperate with a TATA box to enhance the promoter activity. Various Inr elements have been described and classified according to sequence homology, for example, the TdT – Inr family, the PBGD – Inr family, the DHFR – Inr family, the ribosomal protein – Inr family, the adeno – associated virus p5 – Inr family etc. Also it has been shown that TFIID is an integral component of the transcription initiation complex from almost all TATA – less promoters studies. It is strongly suggested that recognition of the Irn by universal or multiple Inr binding proteins (ITF) might provide a means by which a transcription competent complex can assemble (18 – 20).

Myosin is one of only three proteins known to convert chemical energy into mechanical work, i.e. to function as a molecular motor. Each myosin molecule is composed of two myosin heavy chains and four myosin light chains. All smooth muscle myosin heavy chain (SMHC) isoforms (SM1, SM2 & SMB) so far have been shown to be transcribed

from a single gene by alternative RNA splicing which results in divergences at the carboxyl termini (SM1 & SM2), or at the 25/50 kD junction (SMB). The expression of SMHC gene is strictly controlled in a developmental stage – specific and tissue – specific manner. SM1 and SM2 so far have been found to be specific to vascular and nonvascular smooth muscle cells. Smooth muscle cell proliferation in experimental arteriosclerosis and atherosclerosis was associated with dedifferentiation of the smooth muscle cells toward the embryonic phenotype on the basis of the SMHC expression. These existing SMHC isoforms are important molecular markers in the studies of the process of atherosclerosis as well as phenotypic modulation of arterial smooth muscle cells (21 – 26).

As a result of the characterization of the 5' upstream promoter region of the rabbit SMHC gene and the role of untranslated SMHC Exon 1 involved in the tissue – specific regulation of SMHC promoter activity, surprisingly we have found that this 78 bp TATA–less SMHC Exon 1 on its own possesses promoter activity, in the absence of a 5' upstream core promoter region, and even in the absence of any 5' upstream or 3' downstream sequences. This Exon 1 promoter sequence does not contain any TATA motif or CAAT motif to provide the second promoter capacity as mentioned above in the c–myc gene. This Exon 1 promoter also does not contain any known initiator sequences, and for its promoter activity does not require any 5' upstream region. (As is well known, the initiator element usually starts in the 5' upstream region and encompasses the transcription start site).

To investigate this new type of Exon 1 promoter, we have synthesized a series of mutant oligonucleotides of the SMHC Exon 1 sequences —— Exon 1 (+1–+79), Ex 60 (+1–+61), Ex 40 (+1–+41), and Ex 20 (+1–+21), cloned into four pCAT vectors (pCATbasic, pCATpromoter,

pCATenhancer, and pCATcontrol.; Promega) at either promoter or enhancer positions. These vectors were designed specially for testing promoter or enhancer activity and contain no other eukaryotic promoter or enhancer sequences (pCATbasic), or with SV40 promoter but lacking any eukaryotic enhancer (pCATpromoter), or with SV40 enhancer but lacking any eukaryotic promoter (pCATenhancer) or with both SV40 promoter plus enhancer as the positive control (pCATcontrol). Six differentially positioned pCAT – EXON 1 constructions were used to transfect primary rabbit vascular smooth muscle cells (RSMC). Transient expression assays showed that in the absence of the 5' upstrea region, (including the SMHC core promoter, a strong TATA – promoter which can drive CAT gene expression to reach the equivalent level of pCATcontrol) or any other cis–acting regulatory elements located in the 5' upstream or intron 1 regions of the SMHC gene, the whole SMHC Exon 1 sequence (79 bp) alone is sufficient to function as an alternative promoter, to drive the chloramphenicol acetyltransferase (CAT) reporter gene expression in RSMC cells such that the CAT expression level reached 45 – 50% of pCATcontrol, (SV40 promoter plus enhancer). This SMHC exon 1 sequence functions as an alternative promoter, but not enhancer, as the reverse orientation clone of pCATbasic–Exon 1 p could not drive CAT gene expression. The SMHC exon 1 sequence showed no capacity to interact with SV40 promoter (pCATpromoter–Exon 1, Exon 1–pCATpromoter). Further analysis of the promoter activity region located inside of the SMHC Exon 1 sequence showed that the alternative promoter activity requires the whole exon 1 region. The 5' nested deletion mutants of SMHC exon 1 extending from +1 to +21 (Ex20), +41 (Ex40), +61 (Ex60) regions showed no substantial promoter activity. It seems that the 3' end of the SMHC Exon 1 sequence is essential for this alternative promoter activity. Interestingly, a transcription factor binding motifs search using a GCG program

package showed there to be a number of putative transcription factor binding sites in the SMHC Exon 1 region, including: two AP 2 sites (activator protein 2)(27) located at +12–+19 and +16–+23; three GC–rich/GCF sites (28) located at +11–+17, +61–+67 and +63–+69; one SIF site (c–sis/PDGF induced, activates c–fos gene) (29) at +45–+50; and one E box (bHLH proteins binding concensus site i.e. MyoD family) located at +71–+76. It may be that the E box or GC–rich/GCF binding sites at the 3' end region of the SMHC Exon 1 is critical to this exon 1 promoter activity. To answer this question, we have constructed E box and GCF site mutants of SMHC Exon 1, the CACTTG motif of E box having been changed to GGTTTG (EXMUT 1); and the GCGCC central motif of two GC–rich/GCF motifs having been changed to AATTT, (EXMUT 2). Transient expression studies showed that these single mutations of the 3' end motifs dramatically abolish the Exon 1 promoter activity. This phenomenon distinguishes from all initiators reported so far, and requires the whole length of exon 1 to function as an alternative promoter. This is the first time an untranslated TATA–less first exon has been disclosed to have promoter activity in the absence of 5' upstream core promoter sequences. The essential element for its promoter activity is located at the 3' end of the Exon 1 sequence, but not encompassing the transcription start site as defined for the initiator.

From the view of the multifunctions of highly conserved untranslated first exons, it seems these actively transcribed but untranslated sequences are particularly important regions involved in the transcription regulatory mechanisms and pathways which have still to be completely understood. It has been shown that these untranslated first exons can interact with 5' upstream TATA – promoter or other regulatory elements to regulate expression levels or tissue–specificity. It has been

shown that these untranslated first exons can contain a second TATA–promoter which confers an alternative transcription start site. or can contain partial sequences of initiator which encompass the transcription start site.

As another new function of the untranslated first exons of eukaryotic genes, under certain circumstances, for example, in the absence of any 5' upstream core promoter and other regulatory element sequences as well as any 3' flanking sequences, the untranslated first exons can function as an alternative promoter *per se*, even without the necessity of the presence of a potential TATA – box. The transcription initiation complex might be assembled either via specific exon 1 binding proteins, or via known general transcription factors, for instance TBP, either directly binding, or indirectly binding through certain potential tissue–specific and gene–specific transcription factors, as we show here that the MyoD and GC–rich/GCF motifs play a critical role in the SMHC exon 1 promoter activity.

To extend this discovery of a new function of untranslated first exons to other eukaryotic genes, we have also synthesized a 61 nt oligonucleotide containing the whole length of chicken skeletal Troponin I exon 1 sequences and cloned this into pCATbasic vector, to test its promoter activity in mammalian cells. The results show that this untranslated TATA–less first exon of sTnI gene also confers moderate promoter activity in the absence of any 5' upstream sequences including the sTnI core promoter region and without 3' flanking (sTnI intron 1) sequences as well.

SUMMARY OF THE INVENTION

As the basis for this invention, we have discovered a new function of untranslated first exons of eukaryotic genes, without any 5' or 3' flanking sequences namely that they function as an alternative promoter per se, under certain circumstances, for instance, in the absence of 5' upstream normal core promoter sequences.

This exon 1 promoter function has been demonstrated by the synthesized 79 nt SMHC Exon 1 and 61 nt sTnI Exon 1 oligonucleotides (both lacking any 5' or 3' flanking sequences) driving CAT reporter gene expression in mammalian cells to a moderate level —— equivalent to 55 – 60% CAT activity driven by 5' upstream SMHC core promoter or about 45–50% CAT activity driven by SV40 promoter plus enhancer (pCATcontrol).

To dissect the essential elements in the SMHC exon 1 promoter a series of 3' end deletion mutants of the SMHC Exon 1 sequence have been made and transient expression assays showed that the 3' end 18 nucleotides were critical for this exon 1 promoter activity. Further mutations of the MyoD motif or GC–rich motif which are located in this 18 nucleotide 3' end region abolished this exon 1 promoter activity. As is well known, the MyoD motif CANNTG and GC–rich motif GCGCC themselves could not function as a promoter. Hence in one aspect these protein binding motifs located inside the first exon can be utilized as important components of transcription initiation complex assembled by the exon 1 promoter.

This strategy can be an alternative way for highly conserved genes, particularly housekeeping genes, to maintain their basal levels of transcription under certain circumstances, for instance, during genomic rearrangements or gene translocations. The normal gene–specific

functional 5' promoter regions are truncated or inactivated, then the transcribed first exon becomes the potential initiator which requires the whole genetic information to be carried in the gene-specific Exon 1, to direct the temperarly expression of these genes.

In one aspect, the present invention provides the use of SMHC exon 1 promoter to construct moderate expression vectors, for use in Melody Gene Therapy —— the multiple genes regulation strategy in gene therapy, fine controlling of expression levels of concerted genes, and to restore the proper balance genes interactions. The advantages of highly conserved exon 1 promoter are its genetic stability, small size, sequence information always being available, easy manipulation and usually compact in genetic information. The accumulation of exon 1 promoter information provides a new source for the construction of new expression vectors by combining varied motifs, which have been naturally organized inside the first exons to earn some gene-specific character, such as tissue-specific and differentiation – stage specificities.

Furthermore, the invention provides the use of eukaryotic genes first exons sequences in the isolation and characterization of potential transcription factors involved in the initiation and regulation of transcription, as well as in the elucidation of interactions between general and gene-specific transcription factors with eukaryotic polymerase enzymes.

A further aspect of the application of the present invention is as a diagnostic tool in the *in vitro* testing for the response of a patient to gene therapy targeted to inhibit vascular smooth muscle cell proliferation and phenotype change in atherosclerosis and in intimal hyperplasia.

In addition, the SMHC exon 1 promoter can be used as a probe for the identification, isolation, characterization and applications of other genes, particularly of the myosin family including myosin heavy chain genes. These in turn can themselves be used for cell specific expression of phenotype markers and the expression of genes involved in the regulation of cell specific functions, like those of proliferation, migration and phenotype determination.

This invention further provides the use of exon 1 promoter sequences from eukaryotic genes to construct eukaryotic expression vectors for the expression of reporter genes or other marker genes for transient or long term expression studies in mammalian cell cultures or *in vivo* animal model studies.

This invention also provides the use of these eukaryotic exon 1 mini-promoters for gene therapy studies, which avoid the safety issues from viral vectors used in clinical trials which have had to be addressed up to now.

Further, this invention provides the use of these exon 1 promoter sequences to be modified by subtraction or addition or modification of sequences for the construction of eukaryotic expression vectors to modulate expression levels driven by these new eukaryotic promoters or with cell type specific expression, that can be provided in any suitable form appropriate to the protocol of administration and/or needs of a patient undergoing gene therapy or as an in vitro diagnosis tool for use in gene therapy.

BRIEF DESCRIPTION OF THE DRAWINGS

FIGURE 1 illustrates the sequence of the SMHC Exon 1 (+1 to +79) oligonucleotides synthesized in the 5' to 3' orientation, and the putative transcription factors binding sites.

FIGURE 2 illustrates the constructions of pCAT – SMHC Exon 1 plasmid vectors, in which the SMHC Exon 1 sequence was inserted into different positions related to the CAT reporter gene, to test its function as promoter, enhancer or cis–regulatory element in mammalian cells.

FIGURE 3 illustrates the promoter activity of pCAT – SMHC Exon 1 constructs. RSMC, HSMC and other cell–types were transfected with these pCAT – SMHC Exon 1 plasmid DNAs, transient expressed CAT activity was measured by liquid scintillation counting (LSC) and the transfection efficiency was normalized by co–transfection of ß–galactosidase activity and protein content assay. Relative CAT activity is expressed as % of pCATcontrol.

FIGURE 4 illustrates the constructions of SMHC Exon 1 Promoter mutants, including its 3' nested deletion mutants and the MyoD, GC–rich/GCF motifs mutants. All synthesized mutant oligonucleotides were cloned into pCATbasic vector upstream of the CAT reporter gene to test their promoter activity in vascular smooth muscle cells.

FIGURE 5 illustrates the relative CAT activity driven by the SMHC Exon 1 Promoter mutants of Figure 4 in RSMC and HSMC cells.

FIGURE 6 illustrates the synthesized oligonucleotides sequence of sTnI Exon 1, the construction of pCATbasic – sTnI EXon plasmid vector and its expression in mammalian cells.

DETAILED DESCRIPTION OF THE INVENTION AND PREFERRED EMBODIMENTS

In relation to Figures 1 – 6 the following materials and methods were used. Materials: [^{14}C] chloramphenicol (40 – 60 mCi/mmol) and [α – 35 S] – dATP were purchased from Amersham. Enzymes used in plasmid construction were obtained from Boehringer Mannheim Biochemicals and Promega. Plasmids vectors pCATbasic, pCATpromoter, pCATenhancer, pCATcontrol and pSV–βGal were from Promega. Bacterial chloramphenicol acetyltransferase (CAT) and acetyl coenzyme A were from Pharmacia Inc.

Cell cultures: rabbit smooth muscle cells (RSMC) and rabbit endothelial cells (RENDO) were obtained from the thoracic aorta of 8 to 10–weeks old rabbits as described (30 – 31). Human aortic or vein smooth muscle cells (HSMC) were prepared by explantation of aortic or vein tissues obtained during cardiovascular surgery as described (32). Human dermal fibroblasts (HDF) were obtained by enzymatic digestion of tissues as described (33). Human Girard heart cells (HGH) were obtained from ICN Biochemicals Ltd. All cells were maintained in DMEM medium supplemented with 10 – 20% fetal calf serum, 2mM L–glutamine, 0.25 ug/ml fungizone, 100 u/ml penicillin and 100ug/ml streptomycin (all from GIBCO), at 37°C and 10% CO_2. However RENDO cells were maintained in M199 medium supplemented with 20% fetal calf serum, 2mM L–glutamine, 0.25 ug/ml fungizone, 100u/ml penicillin and 100u/ml streptomycin, plus 20 ug/ml endothelial cell growth factor and 80 ug/ml heparin, within gelatin pre–coated flasks at 37°C and 7% CO_2.

Sequence computering analysis: The SMHC gene and sTnI gene sequences obtained from Genbank database. Totally 162 transcription factors binding sites (34) were searched using the GCG program package (*via* UNIX system, SEQNET, Daresbury Laboratory, UK).

DNA sequence analysis: Sequence analysis of the recombinant plasmids was performed by dideoxy chain termination methods (35) using Sequenase (USB). All constructions have been confirmed by sequencing.

Plasmids construction: Equal amounts of the synthetic oligonucleotides were kinase treated and purified by phenol extraction – ethanol precipitation. dsDNA fragments were obtained by hybridizing the upper strand and lower strand oligonucleotides mixed in hybridization buffer (10mM Tris–HCL, pH7.5, 150 mM NaCl, 10mM $MgCl_2$) using single cycle of a PCR instrument (3min/97°C, 15min/65°C, 15min/37°C and 15min/24°C). These dsDNA fragments were cloned into pCAT vectors by appropriate restriction enzyme digestion and blunt–end ligation at the desired insert positions as indicated below. After transformation into $DH_5\alpha$ competent cells, selected recombinants were confirmed by restriction enzyme mapping and DNA sequencing. CsCl gradient method (36) was used to purify the plasmid DNAs for transfection studies.

(1) Exon 1 – pCATbasic: SMHC Exon 1 sequence (+1—+79) was inserted into Sal I site, upstream of the CAT gene, at the promoter position. pCATbasic vector contains pUC19 backbone, polylinker site, CAT gene, SV40 small Tantigen (terminal signal) and ampicillin resistance gene. It lacks any eukaryotic promoter and enhancer sequences.

(2) pCATpromoter – Exon 1: SMHC Exon 1 sequence (+1–+79) was inserted into Sal I site, downstream of CAT gene and SV40 terminal signal, at the enhancer position. pCATpromoter vector contains SV40 promoter but lacks any eukaryotic enhancer sequence.

(3) Exon 1 – pCATpromoter: SMHC Exon 1 sequence (+1–+79) was inserted into Bgl II site, upstream of SV40 promoter and CAT gene at the cis–regulatory element position.

(4) Exon 1 – pCATenhancer: SMHC Exon 1 sequence (+1–+79) was inserted into Sal I site, upstream of CAT gene, at the promoter position. pCATenhancer vector lacks any eukaryotic promoter sequence, but contains SV40 enhancer which was positioned downstream of CAT gene and SV40 terminal signal.

(5) Exon 1 – pCATcontrol: SMHC Exon 1 sequence (+1 – +79) was inserted at Bg1 II site, upstream of SV40 promoter and CAT gene, at the cis–regulatory element position. pCATcontrol contains SV40 promoter plus enhancer which can drive a high level of CAT expression in eukaryotic cells.

(6) pCATcontrol – Exon 1: SMHC Exon1 sequence (+1 – +79) was inserted into Sal I site, downstream of SV40 enhancer, at the enhancer position.

(7) Ex 20 – pCATbasic: 3' end deletion mutant ——

partial SMHC Exon 1 sequence (+1 – +21) was inserted into Sal I site of pCATbasic vector, upstream of CAT gene, at the promoter position.

(8) Ex 40 – pCATbasic: 3' end deletion mutant —— partial SMHC Exon 1 sequence (+1 – +41) was inserted into Sal I site of pCATbasic vector, upstream of CAT gene, at the promoter position.

(9) Ex 60 – pCATbasic: 3' end deletion mutant —— partial SMHC Exon 1 sequence (+1 – +61) was inserted into Sal I site of pCATbasic vector, upstream of CAT gene, at the promoter position.

(10) Exon1op – pCATbasic: SMHC Exon 1 sequence (+79 – +1) was inserted into the Sal I site of pCATbasic vector, but in the reverse orientation, at the promoter position.

(11) EXMUT1 – pCATbasic: MyoD motif mutant of SMHC Exon 1 sequence (+1 – +79), its MyoD motif CACTTG (+71 – +76) was replaced by GGTTTG). Mutated SMHC Exon 1 sequence was inserted into Sal I site of pCATbasic vector, upstream of CAT gene, at the promoter position.

(12) EXMUT2 – pCATbasic: GC–rich/CGF motif mutant of

SMHC Exon 1 sequence (+1 – +79), Two of the GC–rich/GCF motifs located at its 3' end GCGCGCC (+60 – +66) & GCGCCCC (+62 – +69) were mutated by change the central GCGCC motif (+62 – +66) into AATTT. The mutated SMHC Exon 1 sequence was inserted into the Sal I site of pCATbasic vector, upstream of the CAT gene, at the promoter position.

(13) EXMUT2op – pCATbasic: The same insert sequence and position as mentioned above, but was cloned in the reverse orientation.

(14) sTnI Exon1 – pCATbasic: sTnI Exon 1 sequence (+1 – +61) was inserted into the Sal I site of pCATbasic vector, upstream of the CAT gene, at the promoter position.

Cell transfection: All cells were transfected by electroporation method (37), using BIO–RAD gene pulser system. 20–40 ug plasmid DNA was electroporated at 260 V/960 uF (RSMC, HSMC, HGH, HDF) or 230 V/960 uF (RENDO) with $\sim 10^6$ cells suspended in 0.5 ml electroporation buffer. After 48–60 hr, the transfected cells were harvested for CAT activity assays. The transfection efficiency was normalized by co–transfection of ß–Gal expression vector (pSV–ßGal).

CAT enzyme activity: CAT activity was determined as described (38) with some modifications. Briefly, cell extracts in 0.25 M Tris–HCl pH8.0 obtained by freeze thaw were heated at 60°C/10min for the inactivation of endogenous acetylase. Aliquotes of 115 ul cell lysate were incubated with 5 ul of ^{14}C–chloramphenicol (Amersham, 0.025 mCi/ml) and 5 ul of n–butyryl coenzyme A (5ug/ml, Sigma) at 37°C for 60 min. The reaction products were extracted with 300 ul mixed xylenes (Aldrich), the xylene phase was back–extracted with 0.25 M Tris–HC1 pH8.0 twice to remove all unreacted chloramphenical, and use for liquid scintillation counting (LSC).

ß–Gal enzyme assay: Co–transfected ß–Gal activity in Cell lysates (without heat inactivation) was determined as previously described (39).

Protein content assay: The protein content of cell lysates was determined using the protein assay kit (BIO–RAD) based on the method of Bradford (40).

REFERENCES:

1) McKnight S.L. and Kingsbury R.: Science 217 (1982) 316 - 324.

2) Nikovits W JR., Mar J.H. and Ordahl C.P. : Mol. Cell Biol. 10 (1990) 3468 - 3482.

3) Sakonju S., Bogenhagen D.F., and Brown D.D. : Cell 19 (1980) 13 -25.

4) Bogenhagen D.F., Sakonju S. and Brown D.D.: Cell 19 (1980) 27 - 35.

5) Galli G., Hofstetter H., and Birnstiel M.: Nature 294 (1981) 626 - 631.

6) Hultmark D., Klemenz R., and gehring W.J.: Cell 44 (1986) 429 - 438.

7) Klemenz R., Hultmark D., and Gehring W.J. : EMBO J. 4 (1985) 2053 - 2060.

8) McGarry T. J. and Lindquist S. : Cell 42 (1985) 903 - 911.

9) Parker C. S., and Topol J. : Cell 36 (1984) 357 - 369.

10) Theill L. E., Wiborg O., and Vuust J. : Mol. Cell. Biol. 7 (1987) 4329 - 4336.

11) Yong J. Q., Remmers E. F., and Marcu K. B. : EMBO J. 5 (1986) 3553 - 3562.

12) Bently D. L., Brown W. L., and Groudine M. : Genes Dev. 3 (1989) 1179 - 1189.

13) Eick D., Polack A., Kofler E., Lenoir G. M., Rickinson A. B., and Bornkamm G. W. : Oncogene 5 (1990) 1397 - 1402.

14) Meulia T., Krumm A., Spencer C., and Groudine M. : Mol. Cell. Biol. 12 (1992) 4590 - 4600.

15) Sharp P. A. : Cell 68 (1992) 819 - 821.

16) Rigby P. W. : Cell 72 (1993) 7 - 10.

17) Hernandez N. :
Genes Dev. 7 (1993) 1291 - 1308.

18) Weis L., and reinberg D. :
FASEB J. 6 (1992) 3300 - 3309.

19) Javahery R., Khachi A., Kiersten L., Zenzie-Gregory B., and Smale S. T. : Mol. Cell. Biol. 14 (1994) 116 - 127.

20) Usheva A., and Shenk T. :
Cell 76 (1994) 1115 - 1121.

21) Rovner A. S., Thompson M. M., Murphy R. A. :
Am. J. Physiol. 19 (1986) C861 - C870.

22) Eddinger T. J., and Murphy R. A. :
Biochemistry 27 (1988) 3807 - 3811.

23) Nagai R., Kuro-o M., Babij P., Periasamy M., :
J. Biol. Chem. 264 (1989) 9734 - 9737.

24) Babij P. :
Nucleic Acids res. 21 (1993) 1467 - 1471.

25) Aikawa M., Sivam P. N., Kuro-o M., Kimna K., nakahara K. I., Takewaki S. I., Ueda M., Yamaguchi H., Yazaki Y., Periasamy M., Nagai R. : Circulation Res. 73 (1993) 1000 - 1012.

26) Kuro-o M., Nagai R., Nakahara K., Katoh H., Tsai R. C., Tsuchimochi H., Yazaki Y., Ohkubo A., Takaku F. :
J. Biol. Chem. 266 (1991) 3768 - 3773.

27) Williams T. and Tjian R. :
Genes Dev. 5 (1991) 670 - 682.

28) Kegeyama R., and Pastan I. :
Cell 59 (1989) 815 - 825.

29) Wagner B. J., Hayes T. E., Hobab C. J., and Cochran B. H. :
EMBO J. 9 (1990) 4477 - 4484.

30) Yau-Young A. O., Shio H., and Fowler S. :
J. Cell. Physiol. 108 (1981) 416 - 423

31) Voyta J. C., Via D. P., Butlerfield C. E., Zetter B. R.
J. Cell. Biol. 99 (1984) 2034 - 2040.

32) Kantbou C., Parry G., Wigelath E., Kakkar V. V., and Demoliou-Mason C. : FEBS Letters 314 (1992) 143 - 148.

33) Jaffe E. A., Nachman R. L., Becker C. G., and Minick C. R. : J. Clin. Invest. 52 (1973) 2745 - 2756.

34) Faisst. S., and meyer S. : Nucleic Acids Res. 20 (1992) 3 - 26.

35) Sanger T., Nickleen S., and Coulson A. R. : Proc. Natl. Acad. Sci.. USA. 74 (1977) 5463 - 5467.

36) Sambrook J., Fritsch E. F., and Maniatis T. : Molecular Cloning. A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. 1989.

37) Golstein S., Fordis C. M., and Howard B. H. : Nucleic Acids Res. 17 (1989) 3959 - 3971.

38) German C. M., Moffat L. F., and Howard B. : Mol. Cell. Biol. 2 (1982) 1044 - 1051.

39) Rosenthal N. : Methods in Enzymology (1987) 152, 704.

40) Bradford M. : Anal. Biochem. 72 (1976) 248.

CLAIMS

1. The nucleotide seuqence from +1 to +79bp (Exon 1) of the rabbit smooth muscle myosin heavy chain gene or the sequence from +1 to +61 bp (Exon 1) of the chicken skeletal Troponin I gene.

2. The nucleotide sequence of claim 1 for use as a promoter in the expression of one or more genes, esepcially reporter genes, in a mammalian host cell.

3. The nucleotide sequence of claim 1 for use in the construction of a eukaryotic expression vector comprising a reporter or marker gene.

4. A eukaryotic expression vector comprising the nucleotide sequence of claim 1, or a portion or mutant thereof.

5. An expression vector according to claim 4, selected from any of these vectors disclosed herein with reference to the accompanying drawings.

6. A plasmid comprising the nucleotide sequence of claim 1, or a portion or mutant thereof.

7. A plasmid according to claim 6, selected from any of those plasmids disclosed herein with reference to the accompanying drawings.

8. A mammalian host cell transfected with a gene comprising the sequence of claim 1, or a portion or mutant thereof, as the promoter therefor.

9. A cell according to claim 8, which is selected from rabbit or human or other vascular smooth muscle cells, human dermal fibroblasts, human Girard heart cells, rabbit skin fibroblasts, rabbit endothelial cells, rabbit kidney epithelial cells or rat skeletal muscle myoblasts.

10. A method of initiating and/or regulating expression levels and/or tissue specificity of a gene in a mammalian host cell by use of a promoter therefor which comprises the nucleotide sequence of claim 1 or a portion or mutant thereof.

11. A method according to claim 10 wherein the host cell is a human cell.

12. A method according to claim 10, wherein the said gene is that which codes for chloramphenicol acetyl transferase (CAT) or ß–

galactosidase or the firefly luciferase gene.

13. A method according to claim 10, wherein the host cell is a rabbit or human or other vascular smooth muscle cell, human dermal fibroblast, human Girard heart cell, rabbit skin fibroblast, rabbit endothelial cell, rabbit kidney epithelial cell or rat skeletal muscle myoblast.

14. Use of the nucleotide sequence of claim 1, or a portion or mutant thereof, as a promoter in the expression of one or more genes, especially one or more reporter genes, in a mammalian host cell.

15. Use of the nucleotide sequence of claim 1, or a portion or mutant thereof, in the construction of a eukaryotic expression vector for the transient or stable expression of a foreign gene in a mammalian host cell.

16. Use of the nucleotide sequence of claim 1, or a portion or mutant thereof, in the construction of a vector comprising one or more marker genes for vascular SMC phenotype diagnosis.

17. Use of the nucleotide sequence of claim 1, or a portion or mutant

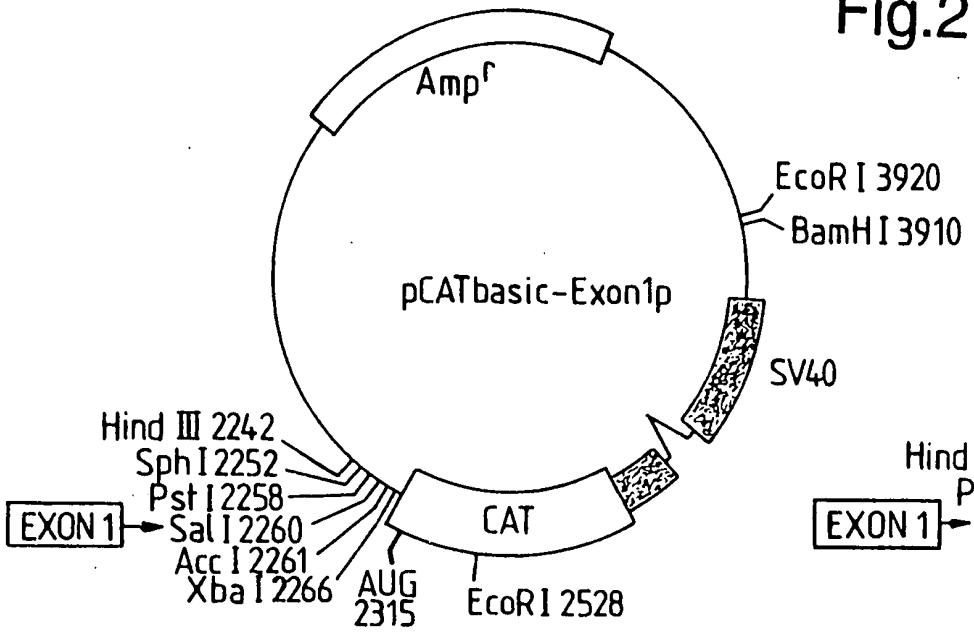
thereof, as a gene probe.

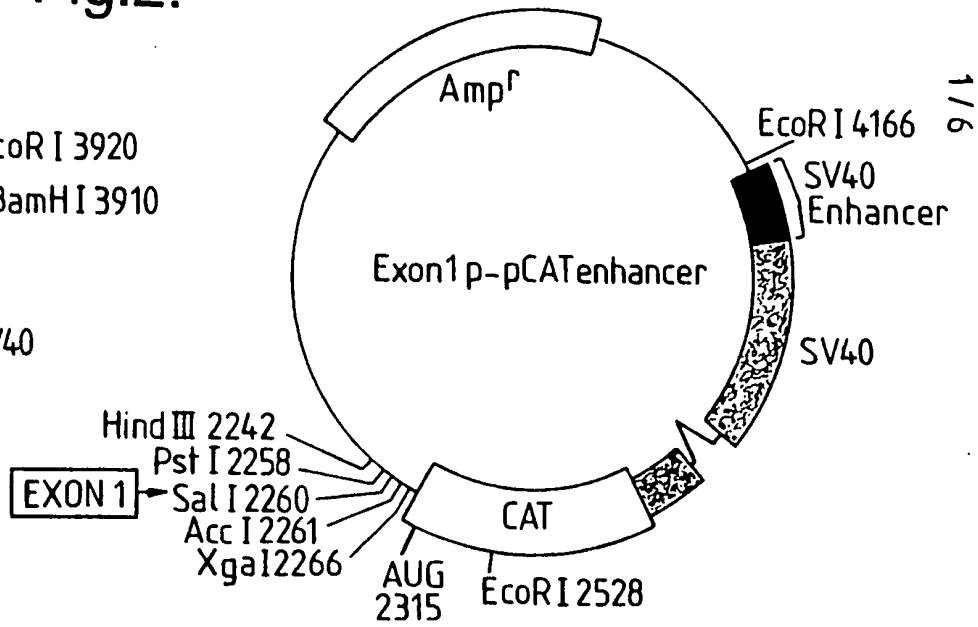
18. Use of the nucleotide sequence of claim 1, or a portion or mutant thereof, in human gene therapy.

Fig.1.

```
1          10      AP2  20        NF1  30
AGCCTTGGACGCCCCGGCCTGGGAGGTGTG
             GC-rich/GCF
31         40          50          60
CCAGACCCGCGCTCCCCGTCCAGTTTCTCC
                  SIF
61         70          79
GCGCGCCCCCACTTGGAG
GCF  GCF  E box
```

Fig.2.





PCT/GB94/01251

Fig.2 (Cont).i

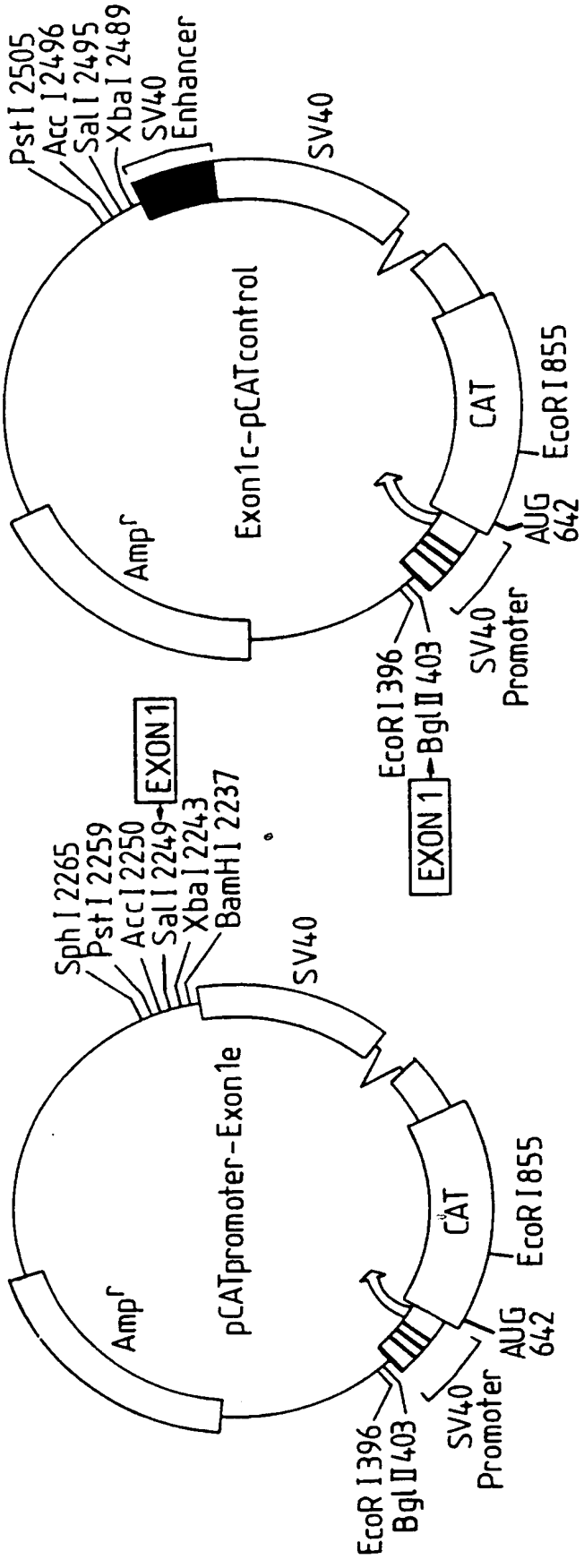
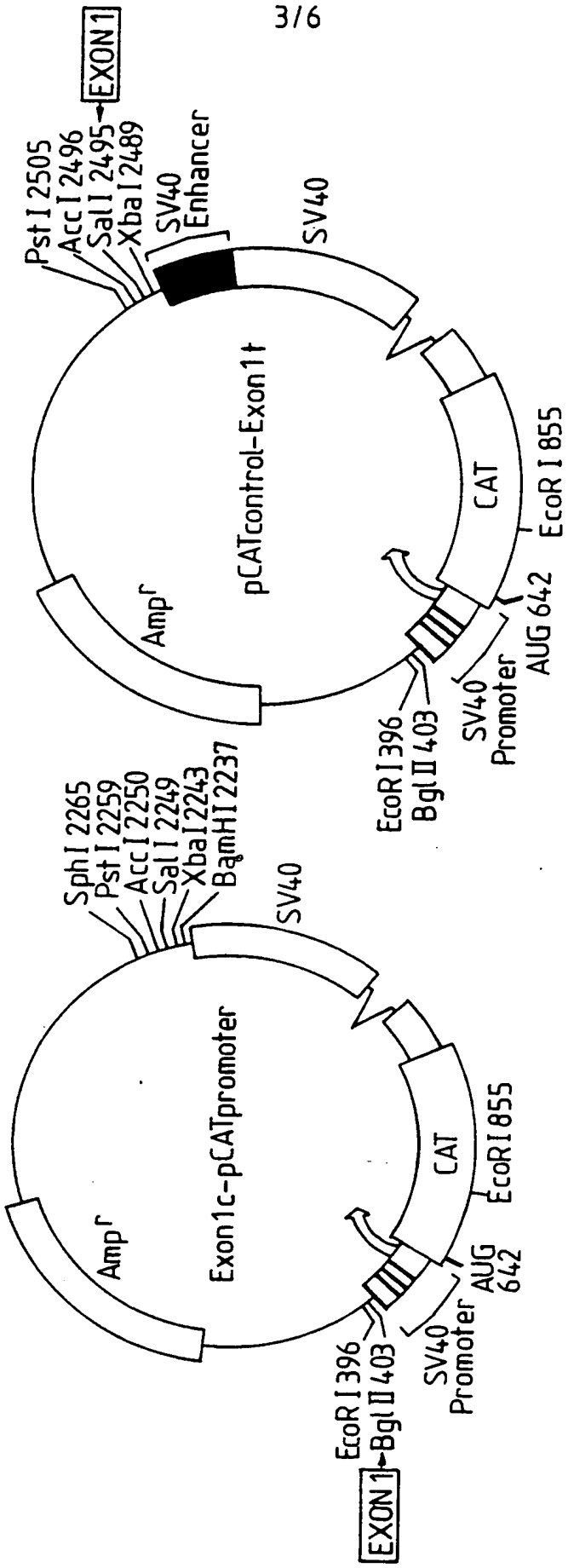


Fig.2 (Cont). ii



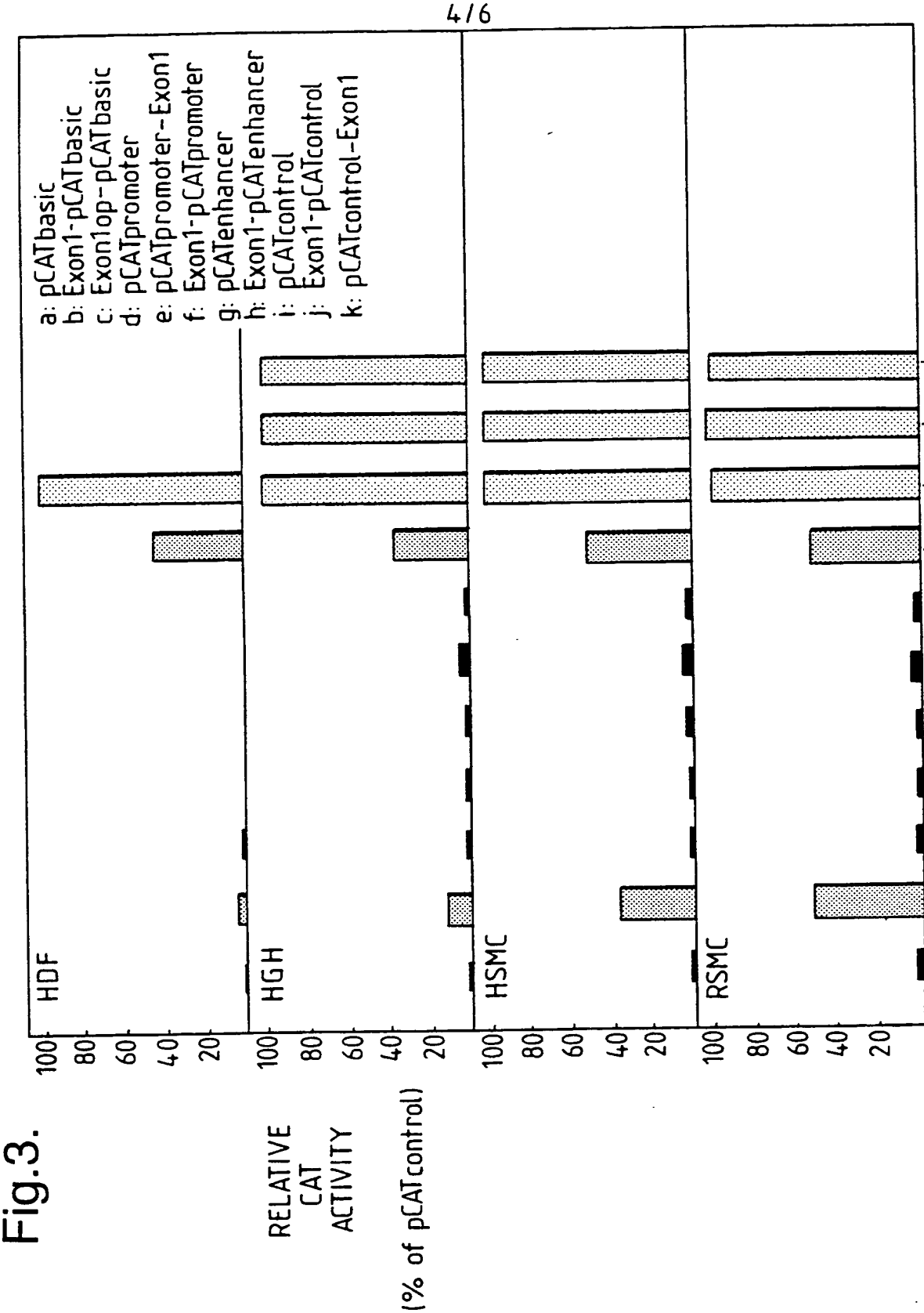
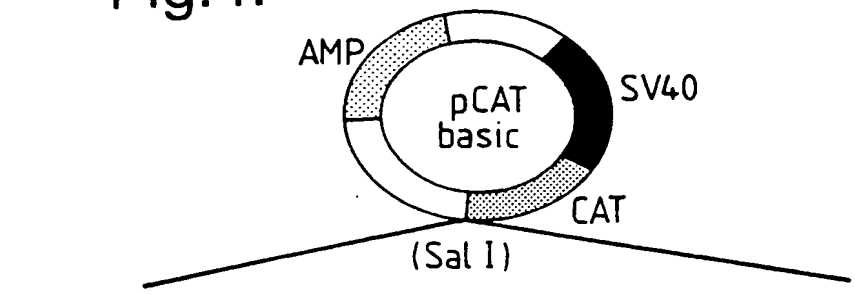


Fig.3.

Fig.4.



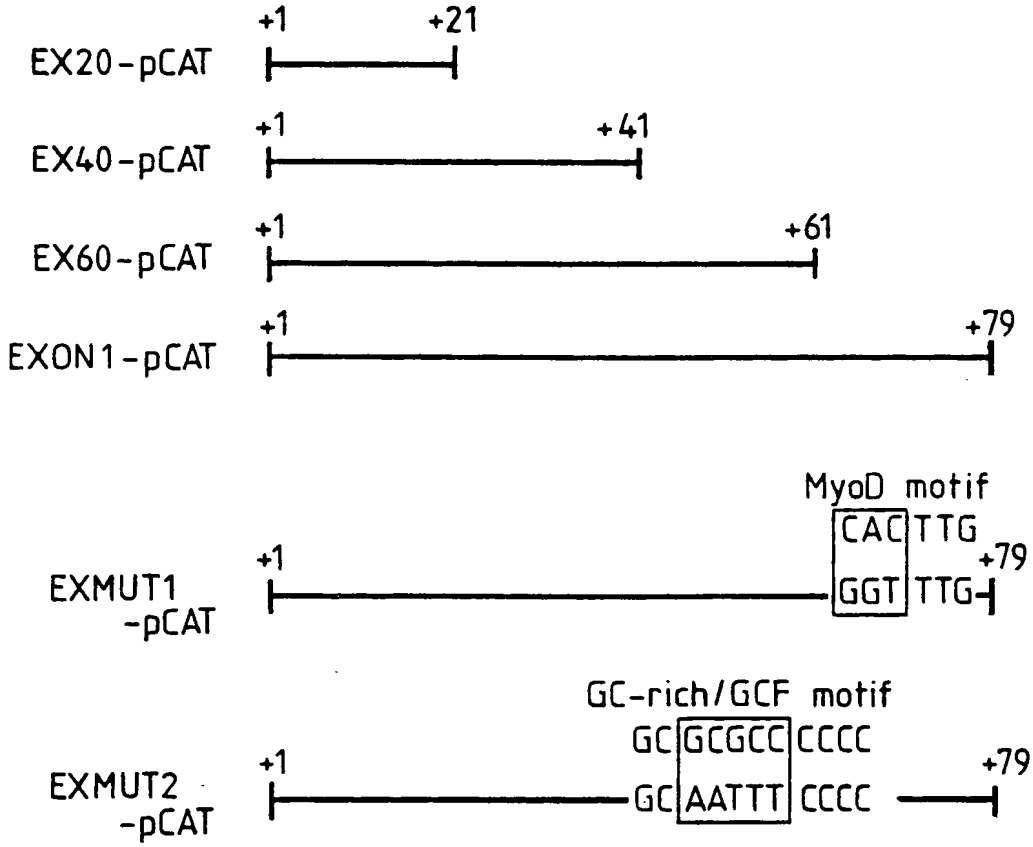


Fig.5.

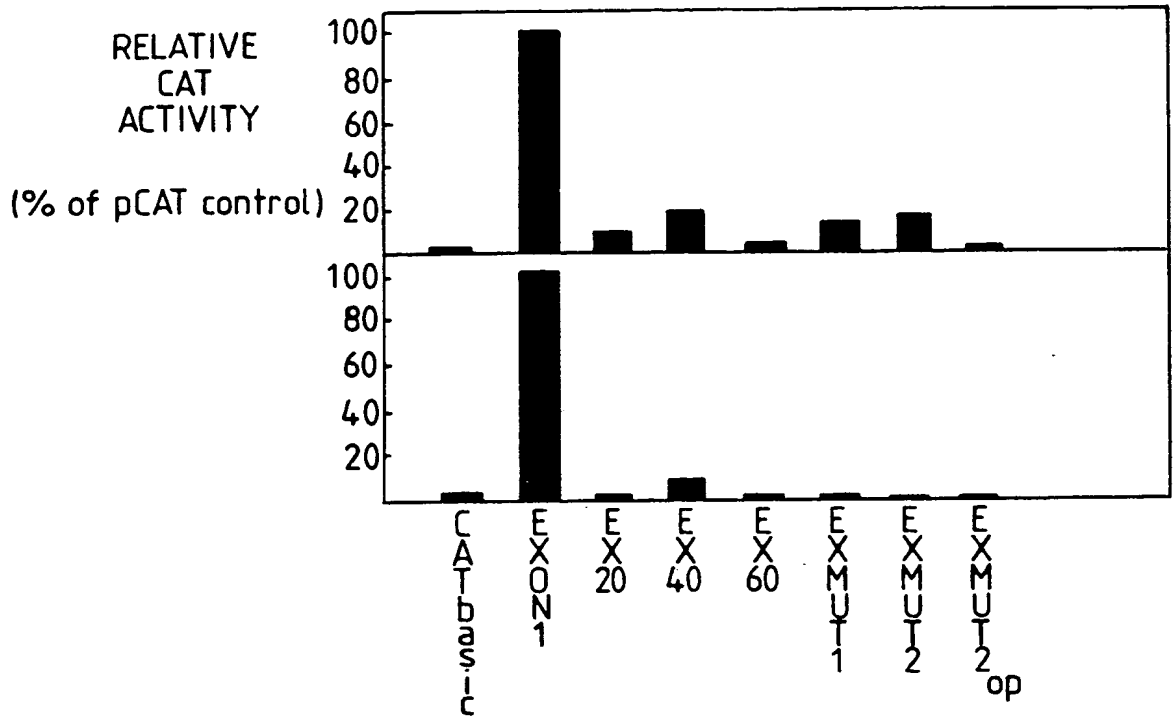
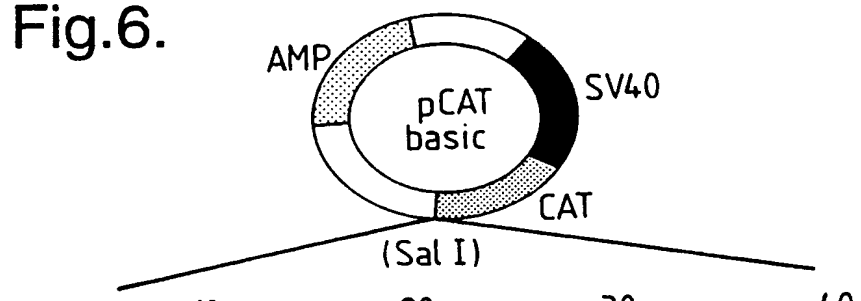
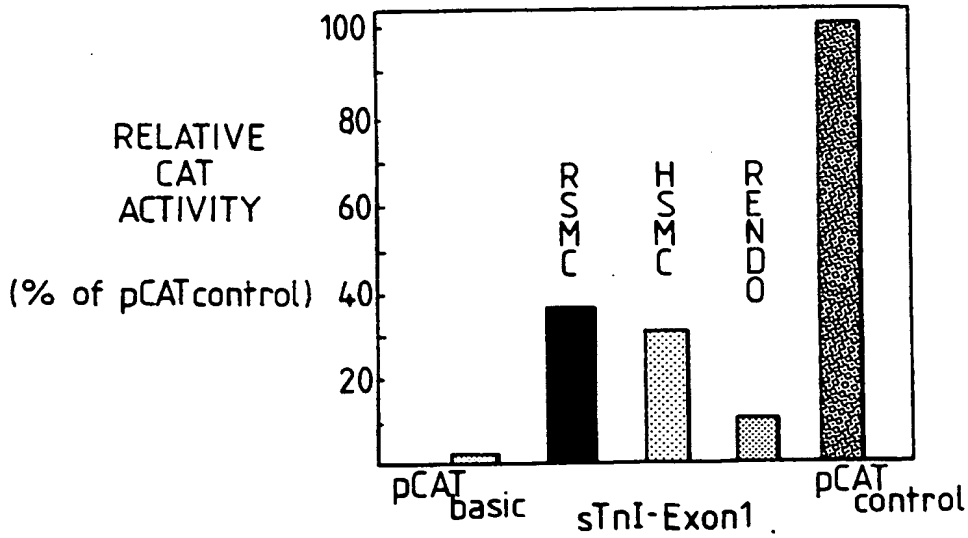


Fig.6.



1 10 20 30 40
TGCGGCTTGCCAGTCAGTCCTGGCCGGCCAGCGGGGCTCT
50 61
CTGTTCCTTCTCCGGACTCGG (sTnI EXON 1)



A. CLASSIFICATION OF SUBJECT MA
IPC 5 C12N15/85 C12N15/12 C12N5/10 C12N9/10 C12Q1/68
A61K48/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 5 C12N C07K A61K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	MOLECULAR AND CELLULAR BIOLOGY, vol.10, no.7, July 1990, WASHINGTON, US pages 3468 - 3482 W. NIKOVITS ET AL 'Muscle-specific activity of the skeletal troponin I promoter requires interaction between upstream regulatory sequences and elements contained within the first transcribed exon' cited in the application see the whole document especially fig 4 and discussion --- -/--	1-15

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

1

Date of the actual completion of the international search	Date of mailing of the international search report
5 October 1994	21. 10. 94

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax (+31-70) 340-3016	Van der Schaal, C

Form PCT/ISA/210 (second sheet) (July 1992)

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA., vol.88, December 1991, WASHINGTON US pages 10676 - 10680 P. BABIJ ET AL 'Characterization of a mammalian smooth muscle myosin heavy-chain gene: complete nucleotide and protein coding sequence and analysis of the 5' end of the gene' see figure 1 see page 10680, left column, last paragraph - right column, paragraph 1 ---	1
X	PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA., vol.82, December 1985, WASHINGTON US pages 8080 - 8084 A. BALDWIN ET AL 'Structure, evolution, and regulation of a fast skeletal muscle troponin I gene' see figure 2 ---	1
E	WO,A,94 20629 (THROMBOSIS RESEARCH INSTITUTE) 15 September 1994 see the whole document especially fig 4 and claims 5-7 -----	1-18

1

INTERNATIONAL [redacted]CH REPORT

Ir. ational application No.
P[redacted]GB94/ 01251

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. [X] Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

remark : Although claim 18 is directed to a method of treatment of the human/animal body, the search has been carried out and based on the alleged effect of the compound.

2. [ ] Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. [ ] Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. [ ] As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. [ ] As all searchable claims could be searches without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. [ ] As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. [ ] No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

[ ] The additional search fees were accompanied by the applicant's protest.

[ ] No protest accompanied the payment of additional search fees.

Inter. nal Application No
PCT/GB 94/01251

Patent docum cited in search report	Publication date	Patent family member(s)	Publication date
WO-A-9420629	15-09-94	NONE	